

## Supplementary material and methods

**Table S1** Sequences and CGI features on each chromosome in the dog genome

Chr.	Chromosome				CGIs				
	# Genes	Length (Mb)	GC content (%)	Obs <sub>CpG</sub> /Exp <sub>CpG</sub>	# CGIs	Length (bp)	GC content (%)	Obs <sub>CpG</sub> /Exp <sub>CpG</sub>	CGI density /Mb
1	1105	121.6	41.5	0.268	3636	1065	61.9	0.744	29.9
2	750	84.3	42.8	0.288	3309	1063	61.7	0.743	39.2
3	454	91.2	40.3	0.240	1719	1073	62.0	0.769	18.8
4	544	88.1	40.3	0.236	1690	1060	61.9	0.763	19.2
5	935	88.5	44.2	0.269	2388	1067	63.0	0.753	27.0
6	834	76.9	42.6	0.270	2067	1109	62.6	0.765	26.9
7	615	80.3	40.9	0.238	1530	1082	62.6	0.771	19.1
8	616	73.9	40.5	0.241	1612	1097	62.6	0.761	21.8
9	1040	60.6	45.9	0.284	2367	1092	63.4	0.749	39.1
10	595	68.9	42.7	0.291	2445	1080	62.3	0.748	35.5
11	496	73.8	40.2	0.233	1522	1152	62.0	0.759	20.6
12	547	72.2	38.9	0.225	1218	1195	62.2	0.770	16.9
13	376	62.6	39.9	0.255	1527	1028	62.4	0.745	24.4
14	370	60.6	38.7	0.229	998	1162	61.5	0.773	16.5
15	461	63.7	40.1	0.272	1687	1068	62.3	0.756	26.5
16	428	58.8	40.9	0.272	1594	1104	61.8	0.744	27.1
17	560	64.0	41.7	0.248	1714	1205	61.8	0.750	26.8
18	687	55.2	42.7	0.291	1918	1077	62.3	0.743	34.7
19	180	53.5	38.3	0.230	840	1114	61.3	0.756	15.7
20	888	57.7	44.3	0.269	2128	1152	63.9	0.748	36.9
21	486	50.7	40.1	0.227	829	1066	62.1	0.770	16.3
22	207	61.2	37.9	0.226	863	1049	61.7	0.763	14.1
23	308	52.1	39.8	0.216	826	1158	62.2	0.781	15.9
24	473	47.4	44.4	0.276	1938	1117	62.9	0.735	40.9
25	377	51.1	41.3	0.264	1626	1153	62.7	0.742	31.8
26	491	38.7	45.4	0.282	1292	1170	62.1	0.750	33.4
27	482	45.6	40.0	0.234	918	1087	62.6	0.764	20.1
28	350	40.9	43.5	0.282	1727	1152	62.0	0.740	42.2
29	204	41.6	38.1	0.233	768	1132	61.0	0.767	18.4
30	367	40.0	41.4	0.250	915	1118	62.8	0.775	22.9
31	207	38.8	39.3	0.273	1262	1174	62.3	0.731	32.5
32	217	38.7	36.9	0.196	342	1235	62.1	0.815	8.8
33	200	31.2	39.1	0.258	723	1087	61.8	0.767	23.2
34	222	41.9	40.3	0.259	919	1047	60.9	0.756	21.9
35	213	26.4	41.3	0.267	789	1065	60.5	0.767	29.9
36	172	30.7	38.5	0.269	649	1136	60.7	0.785	21.2
37	213	30.7	40.1	0.275	809	1095	61.6	0.756	26.3
38	210	23.7	40.8	0.278	701	1067	62.1	0.735	29.6
X	835	122.2	40.2	0.239	2522	1066	61.7	0.736	20.6

**Table S2** Comparison of CpG islands in housekeeping, widely expressed, moderately expressed and narrowly expressed genes

	Housekeeping		Widely expressed		Moderately expressed		Narrowly expressed	
	Total	CGI+ <sup>a</sup> (%)	Total	CGI+ <sup>a</sup> (%)	Total	CGI+ <sup>a</sup> (%)	Total	CGI+ <sup>a</sup> (%)
Dog	1257	863 (68.7)	3416	2193 (64.2)	2599	1506 (57.9)	1357	698 (51.4)
Human	1257	1082 (86.1)	3416	2769 (81.1)	2599	1856 (71.4)	1357	823 (60.6)
Mouse	1257	1033 (82.2)	3416	2625 (76.8)	2599	1705 (65.6)	1357	736 (54.2)

For comparison purpose, only those genes that are homologous in the dog, human and mouse genomes were used.

<sup>a</sup> Number of genes having promoter-associated CGIs.

**Table S3** GO terms that are significantly overrepresented in the D- (without promoter-associated CGIs) genes compared to D+ (with promoter-associated CGIs) genes

GO code	GO term description <sup>a</sup>	D+ (%)	D- (%)	<i>P</i> -value <sup>b</sup>
GO:0004872	M: receptor activity	621 (10.3)	927 (22.4)	2.2×10 <sup>-62</sup>
GO:0007165	B: signal transduction	1372 (22.7)	1474 (35.5)	1.1×10 <sup>-45</sup>
GO:0005886	C: plasma membrane	665 (11.0)	715 (17.2)	1.9×10 <sup>-19</sup>
GO:0005102	M: receptor binding	229 (3.8)	308 (7.4)	9.3×10 <sup>-16</sup>
GO:0005509	M: calcium ion binding	401 (6.6)	461 (11.1)	1.8×10 <sup>-15</sup>
GO:0006811	B: ion transport	495 (8.2)	479 (11.5)	1.7×10 <sup>-8</sup>
GO:0008233	M: peptidase activity	346 (5.7)	334 (8.1)	4.3×10 <sup>-6</sup>
GO:0005578	C: proteinaceous extracellular matrix	119 (2.0)	128 (3.1)	4.0×10 <sup>-4</sup>
GO:0003779	M: actin binding	117 (1.9)	125 (3.0)	5.6×10 <sup>-4</sup>
GO:0006629	B: lipid metabolic process	303 (5.0)	274 (6.6)	7.2×10 <sup>-4</sup>
GO:0005216	M: ion channel activity	262 (4.3)	239 (5.8)	1.2×10 <sup>-3</sup>
GO:0004672	M: protein kinase activity	410 (6.8)	341 (8.2)	7.0×10 <sup>-3</sup>
GO:0004721	M: phosphoprotein phosphatase activity	104 (1.7)	103 (2.5)	9.0×10 <sup>-3</sup>

<sup>a</sup> Gene ontology organizing principles: Cellular component (C), biological process (B) and molecular function (M). Only those GO terms at the 4<sup>th</sup> level or lower were examined.

<sup>b</sup> *P* values were calculated by  $\chi^2$  test for 2×2 contingency table and only the *P*-values < 0.01 were used.

**Table S4** Genes having unknown nucleotides (Ns) in the 2-kb promoter regions

	Dog (%)	Human (%)	Mouse (%)
Genes with $\geq 1$ “N” in the 2-kb sequence	2469	36	25
Genes having promoter-associate CGI(s)	2084 (84.4)	6 (16.7)	6 (24.0)
Genes with $> 20$ “Ns” in the 2-kb sequence	2084	31	24
Genes having promoter-associate CGI(s)	1764 (84.6)	2 (6.5)	5 (20.8)

**Table S5** Comparison of CpG islands (CGIs) in the dog-human-mouse homologous genes after excluding genes having >20 “Ns” in the sequences

	Homologous genes		Essential genes		Non-essential genes	
	Total	CGI+ <sup>a</sup> (%)	Total	CGI+ <sup>a</sup> (%)	Total	CGI+ <sup>a</sup> (%)
Dog	8870	4933 (55.6)	1086	645 (59.4)	894	381 (42.6)
Human	8870	6311 (71.1)	1086	846 (80.0)	894	491 (54.9)
Mouse	8870	5836 (65.8)	1086	772 (71.1)	894	443 (49.6)

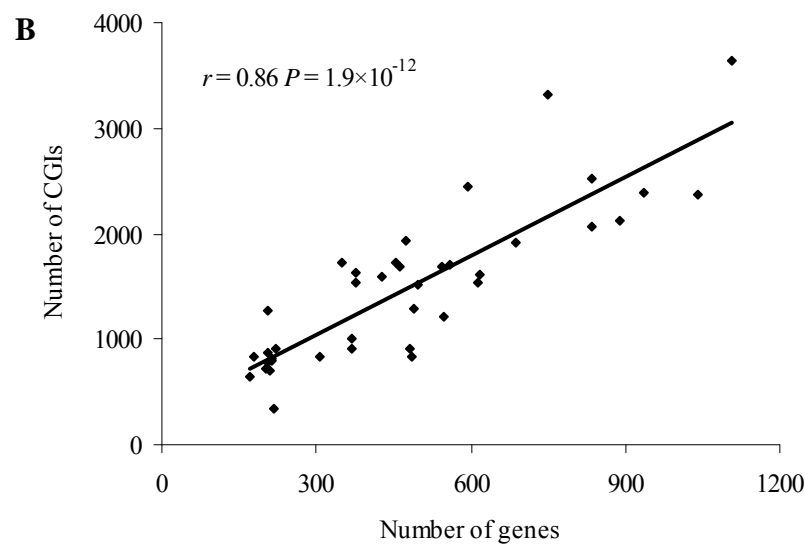
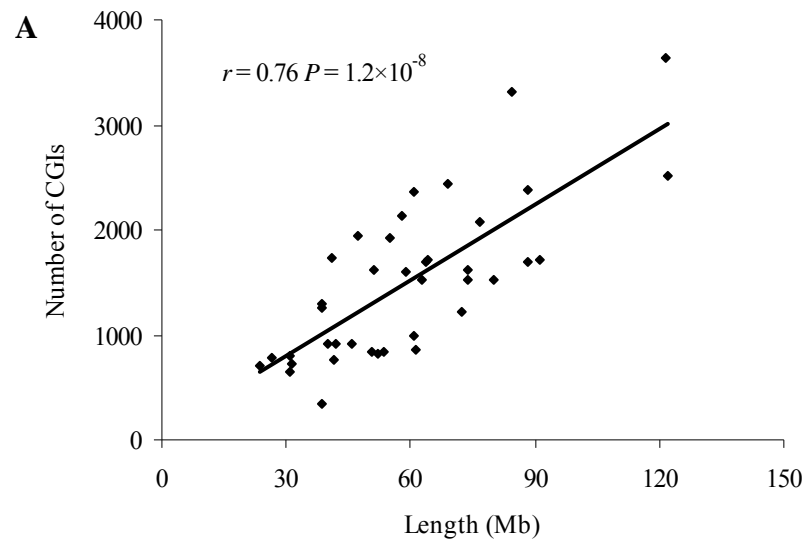
<sup>a</sup> Number of genes having promoter-associated CGIs.

**Table S6** CpG cluster density (/Mb) in genomic regions of three species

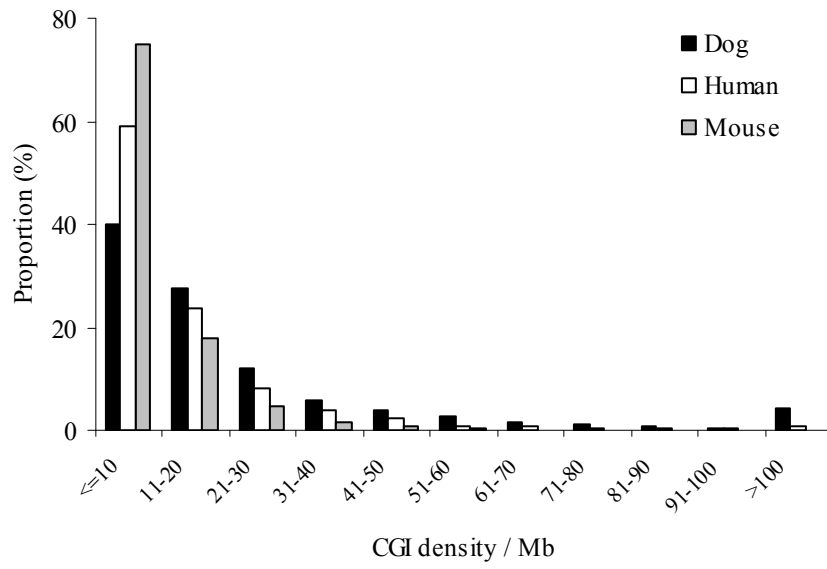
	Genome	Promoter	3' regions	Genic	Intronic	Intergenic
Dog	128.3	674.2	321.1	126.5	109.0	94.1
Human	69.4	501.5	162.1	71.8	56.0	39.6
Mouse	44.4	286.7	84.4	48.0	29.6	26.9

CpG clusters were identified by the CpGcluster algorithm.

**Fig. S1.** Correlation between number of CGIs and genomic factors on each chromosome in the dog genome. (A) Number of CGIs versus chromosome length. (B) Number of CGIs versus number of gene in a chromosome.

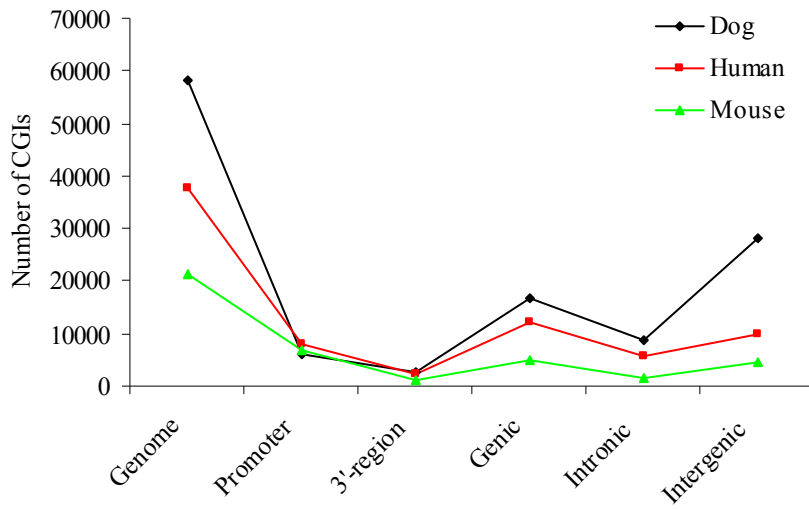


**Fig. S2.** Distribution of CGI density (/Mb) in the dog, human and mouse genomes.





**Fig. S3.** Number of CGIs in the dog, human and mouse genomes and their genomic regions.



**Fig. S4.** Distribution of CGIs in the dog and human homologous genes. A total of 10,196 homologous genes were compared. D+, H+, D-, and H- denote the presence (+) or absence (-) of promoter-associated CGIs in the dog (D) or the human (H) genome.

